UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P O Box 1450
Alexandria, Virginia 22313-1450
www uspto gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/520,922	01/12/2005	Yoshiyuki Okimoto	2004 2068A	4924

513 7590 08/13/2009
WENDEROTH, LIND & PONACK, L.L.P.
1030 15th Street, N.W.,
Suite 400 East
Washington, DC 20005-1503

EXAMINER	
COLUCCI, MICHAEL C	

ART UNIT	PAPER NUMBER
2626	

MAIL DATE	DELIVERY MODE
08/13/2009	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary	Application No.	Applicant(s)	
	10/520,922	OKIMOTO ET AL.	
	Examiner	**Art Unit**	
	MICHAEL C. COLUCCI	2626	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE *3* MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *05/13/2009*.

2a)☒ This action is **FINAL**. 2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-35* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-35* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☒ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

Response to Arguments

1. Applicant's arguments filed 05/13/2009 have been fully considered but they are not persuasive.

Argument (pages 23-24):

- "In the Office Action, the Examiner asserts that the confusable set disclosed in Rigazio is the lower-level N-gram because the N-gram model is applied to a unit smaller than a word. However, in Rigazio, the unit of the lower-level N-gram is "a unit smaller than a word," which is an important difference between Rigazio and the present invention. More specifically, in the present invention, a word is used as a unit even in the "lower-level N-gram." If the "lower-level N-gram" in Rigazio is used for recognizing a title such as "Red Cliff," for example, the sequence is modeled as "r-e- d-c-l-i-f-f." On the other hand, in the present invention, the sequence of the words is modeled, such as "red-cliff," and thus the constraints of the model are stronger than the model in Rigazio" – (page 23 paragraph 5)
- "Conversely, the present invention (as recited in independent claims 1, 13, 14, 26, 27, 29 and 30) models a word string as a title so as to recognize the title even when, for example, the title "Red Cliff" appears only once in training data. On the other hand, the title would not even be grouped into a

partial word string according to the method in Deligne" – (page 24 paragraph 4)

Response to argument:

NOTE: Examiner would like to remind Applicant of the following:

"USPTO personnel are to give claims their broadest reasonable interpretation in light of the supporting disclosure. In re Morris, 127 F.3d 1048, 1054-55, 44 USPQ2d 1023,1027-28 (Fed. Cir. 1997). Limitations appearing in the specification but not recited in the claim should not be read into the claim. E-Pass Techs., Inc. v. 3Com Corp., 343 F.3d1364, 1369, 67 USPQ2d 1947, 1950 (Fed. Cir. 2003) (claims must be interpreted "in view of the specification" without importing limitations from the specification into the claims unnecessarily). In re Prater, 415 F.2d 1393, 1404-05, 162 USPQ 541, 550-551 (CCPA 1969). See also In re Zletz, 893 F.2d 319, 321-22, 13 USPQ2d 1320, 1322 (Fed. Cir. 1989) ("During patent examination the pending claims must be interpreted as broadly as their terms reasonably allow.... The reason is simply that during patent prosecution when claims can be amended, ambiguities should be recognized, scope and breadth of language explored, and clarification imposed.... An essential purpose of patent examination is to fashion claims that are precise, clear, correct, and unambiguous. Only in this way can uncertainties of claim scope be removed, as much as possible, during the administrative process.").

Where an explicit definition is provided by the applicant for a term, that definition will control interpretation of the term as it is used in the claim. Toro Co. v. White Consolidated Industries Inc., 199 F.3d 1295, 1301, 53 USPQ2d 1065, 1069 (Fed. Cir. 1999) (meaning of words used in a claim is not construed in a "lexicographic vacuum, but in the context of the specification and drawings."). Any special meaning assigned to a term "must be sufficiently clear in the specification that any departure from common usage would be so understood by a person of experience in the field of the invention." Multiform Desiccants Inc. v. Medzam Ltd., 133 F.3d 1473, 1477, 45 USPQ2d 1429, 1432 (Fed. Cir. 1998). See also MPEP § 2111.01."

Consider Rigazio's teaching alone, of a well known method of dictionary and syntactic analysis, wherein Rigazio teaches past attempts at improving the recognizer's discrimination among confusingly similar sounds have focused upon the acoustic level 30 and the syntax level 36. At the syntactic level, the <u>erroneous substitution of a confusingly similar sounding word can sometimes be trapped by the syntax rules for word concatenation</u>. For example, in the following two sentences, the acoustic <u>confusability between the words "ate" and "eight" can be discriminated at the syntactic level:</u>

John <u>ate</u> lunch.

John <u>eight</u> lunch.

Similarly, the syntactic level would be able to discriminate between confusably similar words that do not have the identical pronunciation:

John ate lunch.

John nape lunch.

For example:

J . . . O . . . N . . . E . . . S

To further illustrate, FIG. 3 shows how a conventional recognizer adapted for spelled name recognition would deal with the problem of confusable letters. The recognizer 44 employs a language model 46 that is defined based upon all letters of the alphabet (e.g. 26 letters for the English language). The recognizer outputs a sequence or string of letters to the dictionary matching module 48. The dictionary matching module enforces the syntactic rules. It thus discriminates between letter string sequences that are syntactically correct (i.e. that spell names defined by the system) and those that are not. As illustrated diagrammatically at 50, the dictionary matching data structure can be enhanced to include possible confusable letter substitutions, allowing the dictionary to identify a valid name even if one or more of the letters has been replaced by a confusing similarity. Of course, this approach would either greatly increase the

size of the dictionary needed to represent the "syntactically correct" names in their various permutations of misspellings or require a post processing phase (Rigazio Col. 4 line 55 – Col. 3 line 41).

To further improve the well known teachings of Rigazio, Examiner has introduced Deligne in view of Millett, wherein Millett in particular teaches a word in another word, where word streams 44 comprise a plurality of word numbers 56, each of which represent parent and child words 39. As used herein, the phrase "child word" means a word which is related to, describes or comprises additional information about another word (i.e., the parent word). For example, a child word can be a linguistic root of another word (e.g., "peach" is a linguistic root of "peaches"), a sub word of another word (e.g., "CAD" is a sub word of "CAD/CAM"), or a phonetic representation of a word (e.g., "wal'rus" for "walrus"). Illustrated directly below the sentences 36 and 38 are the child words "hunt" and "chase" which, while not expressly part of the sentences 36 and 38, are root words of the parent words "hunted" and "chased", respectively. The words "hunted" and "chased", which are contained in the first and second sentences 36 and 38, are referred to herein as "parent words", because they are the words to which the child words relate. As will be appreciated, the parent words are the same as the file words 37, and a parent word can have a plurality of child words. Parent words are associated with parent nodes 44 of the word list 42 while child words are associated with child nodes 45. While the child words "hunt" and

"chase" are not literally part of sentences 36 or 38, it should be understood that it is possible for child words to also form parts of sentences, etc., such that a child word can also be a parent word. As shown in FIG. 2, the word number "4" is repeated in the word stream 44A, because the word "a" is in both the first and second sentences 36 and 38. The word stream 44A has a sentence level granularity and, therefore, contains granule markers 58 delineating the beginning and end of the first and second sentences 36 and 38. In contrast, the word stream 44B is a word level granularity with granule markers delineating the beginning and end of each file word 37 and its child words 39 (Millett Col. 5 lines 20-58).

Further, Millett teaches the parent word "banking" has "bank" as a root child word. However, if the root child word "bank" is a noun (e.g., such as in the sense of a financial institution), it would not correspond to the parent word "banking" if this parent word is used in the context of a verb (e.g., as in banking a plane). In this case, while the child word "bank" in the form of a noun is a child word of the parent word "banking", it would not correspond to its parent word. In contrast, the child word "bank" used in the context of a verb would correspond to the parent word "banking" used in the context of a verb. In context sensitive situations (e.g., noun/verb determinations, etc.), the execution of decision block 102 can have to be postponed until the next file word is retrieved in block 82 so that the context of the preceding file word can be

properly evaluated (Millett Col. 7 line 65 – Col. 8 line 12).

Millett also teaches the identification of words and their frequency of occurrence to further assist in proper evaluation, wherein Millett teaches list file (or an Alpha Word List as described in the Millett Patent). Referring to Table 1 below, an exemplary Alpha Word List is illustrated which contains the word (both parent and child alphabetically listed), the word number, the number of granules in which the word occurred (frequency count) and whether the word is a child word for a word level granularity for the first file 30. The above described Alpha Word List is created by visiting each element 146 of the element table 144 (FIG. 9). Within each element 146, the binary trees under the sub-elements 152 are traversed and merged in alphabetical order. The information for each word is then written to the Alpha Word List file as the word list 142 is traversed. While traversing each entry keep statistics from the frequency counts to calculate memory needs for Phase II processing (Millett Col. 11 line 39 – Col. 12 line 29 & Table 1).

The teachings of Millett clearly establish an obvious improvement of sentential parsing with respect to word analysis in relation to various parts of speech (verb, noun, etc.) as well as context, wherein the differentiation of child and parent words (i.e. a word as part of another word) allows for an improved evaluation of the speech input of Rigazio and Deligne.

Claim Rejections - 35 USC § 103

2. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

3. Claims 1-5, 7, 9, 13-18, 23 and 26-35 rejected under 35 U.S.C. 103(a) as being unpatentable over Rigazio et al. US 6182039 B1 (hereinafter Rigazio) in view of Deligne et al. US 6314399 B1 (hereinafter Deligne) and further in view of Millett et al. US 6584458 B1 (hereinafter Millett).

Re claims 1, 13, 14, and 26-30, Rigazio teaches language model generation and accumulation apparatus that generates and accumulates language models for speech recognition, the apparatus comprising:

a lower-level N-gram language model (Col. 6 lines 11-20) generation and accumulation unit operable to generate and accumulate a lower-level N-gram language model that is obtained by modeling (Col. 4 lines 30-55 & Fig. 2) a sequence of two or more words within the word string class;

an alignment of words is recognized from an input speech, by referring to a recognition dictionary which describes pronunciation of the words Col. 4 line 55 – Col. 3 line 41),

However, Rigazio fails to teach a word string class and a plurality of text as a second sequence of words that includes the word string class

a higher-level N-gram language model generation and accumulation unit operable to generate and accumulate a higher-lever N-gram language model that is obtained by modeling each of a plurality of texts as a sequence of words that includes a word string class indicating a linguistic property of a word string constituting two or more words and (ii) at least one word included in the plurality of texts except for the words included in the word string class;

a sequence of words including the word string class is assumed in the alignment of' words,

Deligne teaches well known limitations of previous technology, wherein Deligne teaches class versions of phrase based models can be defined in a way similar to the way class version of N-gram models are defined, i.e., by assigning class labels to the phrases. In prior art it consists in first assigning word class labels to the words, and in then defining a phrase class label for each distinct phrase of word class labels. A drawback of this approach is that only phrases of the same length can be assigned the same class label. For example, the phrases "thank you" and "thank you very much" cannot be assigned the same class label, because being of different lengths, they will lead to different sequences of word class labels (Deligne Col. 2 lines 10-20).

Further, Deligne improves these limitations by teaching the clustering (classification process) of the variable-length phrases is explained. Recently, class-

phrase based models have gained some attention, but usually like in Prior Art Reference 1, it assumes a previous clustering of the words. Typically, each word is first assigned a word-class label C.sub.k, then variable-length phrases, wherein the phrases "thank you for" and "thank you very much for" cannot be assigned the same class label. In the present preferred embodiment, it is proposed to address this limitation by directly clustering phrases instead of words (Deligne Col. 10 lines 43-60)

Furthermore, Deligne teaches the step ensures that the class assignment based on the mutual information criterion is optimal with respect to the current phrase distribution, and the step SS2 ensures that the bigram distribution of the phrases optimizes the likelihood calculated according to Equation (19) with the current class distribution. The training data are thus iteratively structured at a both paradigmatic and syntagmatic level in a fully integrated way (the terms paradigmatic and syntagmatic are both linguistic terms). That is, the paradigmatic relations between the phrases expressed by the class assignment influence the reestimation of the bigram distribution of the phrases, while the bigram distribution of the phrases determines the subsequent class assignment (Deligne Col. 11 lines 29-43).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Rigazio to incorporate a word string class and a plurality of text as a second sequence of words that includes the word string class and a higher-level N-gram language model generation and accumulation unit operable to generate and accumulate a higher-lever N-gram language model that is obtained by modeling each of a plurality of texts as a sequence of words that includes a word string

class indicating a linguistic property of a word string constituting two or more words as taught by Deligne to allow for optimal class assignment to account for sentence and word based modeling in speech recognition (Deligne Col. 10 lines 43-60).

Additionally, Deligne teaches a deterministic model, where there is no ambiguity on the parse of a sentence into phrases, whereas in a stochastic model various ways of parsing a sentence into phrases remain possible. For this reason, stochastic models can be expected to evidence better generalization capabilities than deterministic models. For example, assuming that the sequence [bcd] is in the inventory of sequences of the model, then, in the context of a deterministic model, the string "b c d" will be parsed as being a single sequence "[bcd]". On the other hand, in the context of a stochastic model, the possibility of parsing the string "b c d" as "[b] [c] [d] ", "[b] [cd]" or "[bc] [d]" also remain. Class versions of phrase based models can be defined in a way similar to the way class version of N-gram models are defined, i.e., by assigning class labels to the phrases. In prior art it consists in first assigning word class labels to the words, and in then defining a phrase class label for each distinct phrase of word class labels (Deligne Col. 2 lines 1-24).

Deligne also teaches a statistical class sequence model called A class bi-multigram model from input training strings of discrete-valued units, where bigram dependencies are assumed between adjacent variable length sequences of maximum length N units, and where class labels are assigned to the sequences. The number of times all sequences of units occur are counted, as well as the number of times all pairs

of sequences of units co-occur in the input training strings. An initial bigram probability distribution of all the pairs of sequences is computed as the number of times the two sequences co-occur, divided by the number of times the first sequence occurs in the input training string. Then, the input sequences are classified into a pre-specified desired number of classes. Further, an estimate of the bigram probability distribution of the sequences is calculated by using an EM algorithm to maximize the likelihood of the input training string computed with the input probability distributions. The above processes are then iteratively performed to generate statistical class sequence model (Deligne Abstract & Fig. 1).

Therefore, it would have also been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Rigazio to incorporate at least one word included in the plurality of texts except for the words included in the word string class and a sequence of words including the word string class is assumed in the alignment of' words as taught by Deligne to allow for various classes of sentences, wherein sequences of words within a sentence represent distinct class labels, whereby frequency counts associated with said class labels are isolated from the rest of the parsed words (Deligne Col. 2 lines 1-24).

However, Rigazio in view of Deligne fails to teach a word string class that further includes a virtual word denoting a beginning of the word string class and a virtual word denoting and end of the word string class.

the higher-level N-gram language model is an N-gram language model for calculating a link between the words or a word that can be broken down into a plurality of words.

the lower-level N-~ram language model is an N-~ram language model for calculating a

link between the words included in the word that can be broken down into the plurality of words, in the speech recognition.

the input speech is recognized based on (i) a probability that tile words including the word string class appear in an order of appearance in the assumed sequence of words and (ii) a probability of an appearance of the words or the virtual word denoting the end of the word string class in an order of appearance in the word string class

Millett teaches that it is necessary to increase the current `virtual word number` up to the beginning of the next word cluster boundary whenever the end of a data item is reached in the word stream. The only way to know this is to place a marker in the word stream signaling the end of a data item. For non-word level indexes, granules naturally fall within data items, so there is not a problem with a row in the granule cross reference table referring to more than one data item (Millett Col. 11 lines 19-29).

Further, Millett teaches that granule boundary markers 58 are used to demarcate the beginning and end of granules (e.g., "<MB>" for the beginning of a granule and "<ME>" for the end of a granule 60), as shown in FIG. 2. As used herein, the term "granule" and its derivatives refers to a predetermined set of text, or an indexing unit.

The granule size determines the degree to which the location of a word within a document can be determined. For example, a document level granularity would be able to identify the document in which a word appears but not the page or paragraph. A paragraph level granularity would-be able to more precisely identify the paragraph within a document where a word appears, while a word level granularity would be able to identify the sequential word location of a word (e.g., the first word of the document, the second word of the document, etc.). As the granularity increases and approaches word level granularity, the size and complexity of an index increases, but word locations can be more precisely defined. The purpose of the word stream 44 is to track the granules in which a word occurs, not the total number of occurrences of the word. (Millett Col. 4 line 50 – Col. 5 line 15).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Rigazio in view of Deligne to incorporate a word string class that further includes a virtual word denoting a beginning of the word string class and a virtual word denoting and end of the word string class as taught by Millett to allow for the identification of varying text (i.e. phrases or words or paragraphs), wherein markers (i.e. virtual words) are used to tag the beginning and end of the specified granule (i.e. phrases, words, paragraphs, etc.) (Millett Col. 4 line 50 – Col. 5 line 15).

Millett teaches word streams 44 comprise a plurality of word numbers 56, each of which represent parent and child words 39. As used herein, the phrase "child word"

means a word which is related to, describes or comprises additional information about another word (i.e., the parent word). For example, a child word can be a linguistic root of another word (e.g., "peach" is a linguistic root of "peaches"), a sub word of another word (e.g., "CAD" is a sub word of "CAD/CAM"), or a phonetic representation of a word (e.g., "wal'rus" for "walrus"). Illustrated directly below the sentences 36 and 38 are the child words "hunt" and "chase" which, while not expressly part of the sentences 36 and 38, are root words of the parent words "hunted" and "chased", respectively. The words "hunted" and "chased", which are contained in the first and second sentences 36 and 38, are referred to herein as "parent words", because they are the words to which the child words relate. As will be appreciated, the parent words are the same as the file words 37, and a parent word can have a plurality of child words. Parent words are associated with parent nodes 44 of the word list 42 while child words are associated with child nodes 45. While the child words "hunt" and "chase" are not literally part of sentences 36 or 38, it should be understood that it is possible for child words to also form parts of sentences, etc., such that a child word can also be a parent word. As shown in FIG. 2, the word number "4" is repeated in the word stream 44A, because the word "a" is in both the first and second sentences 36 and 38. The word stream 44A has a sentence level granularity and, therefore, contains granule markers 58 delineating the beginning and end of the first and second sentences 36 and 38. In contrast, the word stream 44B is a word level granularity with granule markers delineating the beginning and end of each file word 37 and its child words 39 (Millett Col. 5 lines 20-58).

Further, Millett teaches the parent word "banking" has "bank" as a root child word. However, if the root child word "bank" is a noun (e.g., such as in the sense of a financial institution), it would not correspond to the parent word "banking" if this parent word is used in the context of a verb (e.g., as in banking a plane). In this case, while the child word "bank" in the form of a noun is a child word of the parent word "banking", it would not correspond to its parent word. In contrast, the child word "bank" used in the context of a verb would correspond to the parent word "banking" used in the context of a verb. In context sensitive situations (e.g., noun/verb determinations, etc.), the execution of decision block 102 can have to be postponed until the next file word is retrieved in block 82 so that the context of the preceding file word can be properly evaluated (Millett Col. 7 line 65 – Col. 8 line 12).

Millett also teaches the identification of words and their frequency of occurrence to further assist in proper evaluation, wherein Millett teaches list file (or an Alpha Word List as described in the Millett Patent). Referring to Table 1 below, an exemplary Alpha Word List is illustrated which contains the word (both parent and child alphabetically listed), the word number, the number of granules in which the word occurred (frequency count) and whether the word is a child word for a word level granularity for the first file 30. The above described Alpha Word List is created by visiting each element 146 of the element table 144 (FIG. 9). Within each element 146, the binary trees under the sub-elements 152 are traversed and merged in alphabetical order. The information for each word is then written to the Alpha Word List file as the word list 142 is traversed. While

traversing each entry keep statistics from the frequency counts to calculate memory needs for Phase II processing (Millett Col. 11 line 39 – Col. 12 line 29 & Table 1).

Therefore, it would have also been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Rigazio in view of Deligne to incorporate the higher-level N-gram language model is an N-gram language model for calculating a link between the words or a word that can be broken down into a plurality of words, and the lower-level N-gram language model is an N-gram language model for calculating a link between the words included in the word that can be broken down into the plurality of words, in the speech recognition, and the input speech is recognized based on (i) a probability that tile words including the word string class appear in an order of appearance in the assumed sequence of words and (ii) a probability of an appearance of the words or the virtual word denoting the end of the word string class in an order of appearance in the word string class as taught by Millett to allow for the identification of varying text (i.e. phrases or words or paragraphs), wherein markers (i.e. virtual words) are used to tag the beginning and end of the specified granule (i.e. phrases, words, paragraphs, etc.) (Millett Col. 4 line 50 – Col. 5 line 15), whereby sub-words are properly identified to allow for proper contextual understanding (Millett Col. 11 line 39 – Col. 12 line 29 & Table 1).

Re claims 2 and 15, Rigazio teaches the language model generation and accumulation apparatus according to Claim 1, wherein the higher-level N-gram language model (Col. 6 lines 11-20) generation and accumulation unit and the lower-

level N-gram language model generation and accumulation unit generate the respective language models (Col. 4 lines 4-55 & Fig. 2), using different corpuses (Col. 7 line 21 – Col. 8 line 19).

However, Ragazio fails to teach the higher-level N-gram language model of claim 1.

Deligne teaches well known limitations of previous technology, wherein Deligne teaches class versions of phrase based models can be defined in a way similar to the way class version of N-gram models are defined, i.e., by assigning class labels to the phrases. In prior art it consists in first assigning word class labels to the words, and in then defining a phrase class label for each distinct phrase of word class labels. A drawback of this approach is that only phrases of the same length can be assigned the same class label. For example, the phrases "thank you" and "thank you very much" cannot be assigned the same class label, because being of different lengths, they will lead to different sequences of word class labels (Deligne Col. 2 lines 10-20).

Further, Deligne improves these limitations by teaching the clustering (classification process) of the variable-length phrases is explained. Recently, class-phrase based models have gained some attention, but usually like in Prior Art Reference 1, it assumes a previous clustering of the words. Typically, each word is first assigned a word-class label C.sub.k, then variable-length phrases, wherein the phrases "thank you for" and "thank you very much for" cannot be assigned the same class label. In the present preferred embodiment, it is proposed to address this limitation by directly clustering phrases instead of words (Deligne Col. 10 lines 43-60)

Furthermore, Deligne teaches the step ensures that the class assignment based on the mutual information criterion is optimal with respect to the current phrase distribution, and the step SS2 ensures that the bigram distribution of the phrases optimizes the likelihood calculated according to Equation (19) with the current class distribution. The training data are thus iteratively structured at a both paradigmatic and syntagmatic level in a fully integrated way (the terms paradigmatic and syntagmatic are both linguistic terms). That is, the paradigmatic relations between the phrases expressed by the class assignment influence the reestimation of the bigram distribution of the phrases, while the bigram distribution of the phrases determines the subsequent class assignment (Deligne Col. 11 lines 29-43).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Rigazio to incorporate higher level language modeling as taught by Deligne to allow for optimal class assignment to account for sentence and word based modeling in speech recognition (Deligne Col. 10 lines 43-60).

Re claims 3 and 16, Rigazio teaches the language model generation and accumulation apparatus according to Claim 2, wherein the lower-level N-gram language model (Col. 6 lines 11-20) generation and accumulation unit includes a corpus update unit operable to update the corpus (Col. 12 lines 23-41) for the lower-level N-gram language model (Col. 4 lines 4-55 & Fig. 2),

the lower-level N-gram language model generation and accumulation unit updates the lower-level N-gram language model based on the updated corpus (Col. 12 lines 23-41), and generates the updated lower-level N-gram language model (Col. 4 lines 4-55 & Fig. 2).

Re claims 4 and 17, language model generation and accumulation apparatus according to Claim 1, wherein the lower-level N-gram language model (Col. 6 lines 11-20) generation and accumulation unit analyzes the first sequence of words (Col. 4 lines 4-55 & Fig. 2), and generates the lower-level N-gram language model by modeling each sequence of the one or more morphemes based on the word string class (Col. 4 lines 4-55 & Fig. 2).

However, Rigazio fails to teach analyzing the first sequence of words within the word string class into one or more morphemes that are the smallest language units having meanings.

Deligne teaches that the N-gram class model is defined as a language model that approximates a word N-gram in combinations of occurrence distributions of word-class N-grams and class-based words as shown by the following equation (this equation becomes equivalent to an HMM equation in morphological or morphemic analysis if word classes are replaced by parts of speech (Deligne Col. 18 lines 1-16).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Rigazio to incorporate analyzing the first sequence of words within the word string class into one or more morphemes that are

the smallest language units having meanings as taught by Deligne to allow for a multidimensional probabilistic method of prediction used to classify and model speech, wherein analysis can be performed on the smallest text units (i.e. morphemes) (Deligne Col. 18 lines 1-16).

Re claims 5 and 18, language model generation and accumulation apparatus according to Claim 1, wherein the higher-level N-gram language model (Col. 6 lines 11-20) generation and accumulation unit, and then generates the higher-level N-gram language model by modeling (Col. 4 lines 30-55 $ Fig. 2)

However, Rigazio fails to teach the word string class being included in each of the plurality of texts analyzed into morphemes

Deligne teaches that the N-gram class model is defined as a language model that approximates a word N-gram in combinations of occurrence distributions of word-class N-grams and class-based words as shown by the following equation (this equation becomes equivalent to an HMM equation in morphological or morphemic analysis if word classes are replaced by parts of speech (Deligne Col. 18 lines 1-16).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Rigazio to incorporate the word string class being included in each of the plurality of texts analyzed into morphemes as taught by Deligne to allow for a multidimensional probabilistic method of prediction used to

classify and model speech, wherein analysis can be performed on the smallest text units (i.e. morphemes) (Deligne Col. 18 lines 1-16).

However, Rigazio in view of Deligne fails to teach a sequence made up of the virtual word and the other words

substituting the word string class with a virtual word

Millett teaches that it is necessary to increase the current `virtual word number` up to the beginning of the next word cluster boundary whenever the end of a data item is reached in the word stream. The only way to know this is to place a marker in the word stream signaling the end of a data item. For non-word level indexes, granules naturally fall within data items, so there is not a problem with a row in the granule cross reference table referring to more than one data item (Millett Col. 11 lines 19-29).

Further, Millett teaches that granule boundary markers 58 are used to demarcate the beginning and end of granules (e.g., "<MB>" for the beginning of a granule and "<ME>" for the end of a granule 60), as shown in FIG. 2. As used herein, the term "granule" and its derivatives refers to a predetermined set of text, or an indexing unit. The granule size determines the degree to which the location of a word within a document can be determined. For example, a document level granularity would be able to identify the document in which a word appears but not the page or paragraph. A paragraph level granularity would-be able to more precisely identify the paragraph within a document where a word appears, while a word level granularity would be able to identify the sequential word location of a word (e.g., the first word of the document,

the second word of the document, etc.). As the granularity increases and approaches word level granularity, the size and complexity of an index increases, but word locations can be more precisely defined. The purpose of the word stream 44 is to track the granules in which a word occurs, not the total number of occurrences of the word. (Millett Col. 4 line 50 – Col. 5 line 15).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Rigazio in view of Deligne to incorporate substituting the word string class with a virtual word as taught by Millett to allow for the identification of varying text (i.e. phrases or words or paragraphs), wherein markers (i.e. virtual words) are used to tag the beginning and end of the specified granule (i.e. phrases, words, paragraphs, etc.) (Millett Col. 4 line 50 – Col. 5 line 15).

Re claims 7, 9, and 22, Rigazio teaches the language model generation and accumulation apparatus according to Claim 1, further comprising

a syntactic tree generation unit operable to perform morphemic analysis as well as syntactic analysis of a text (Col. 5 lines 42-63), and generate a syntactic tree in which said-the text is structured by a plurality of layers, focusing on a node that is on said the syntactic tree (Col. 5 lines 42-63) and that has been selected on the basis of a predetermined criterion (Col. 4 lines 4-55 & Fig. 2),

wherein the higher-level N-gram language model (Col. 6 lines 11-20) generation and accumulation unit generates the higher-level N-gram language model for syntactic

tree, using a first subtree (Col. 5 lines 42-63 & Fig. 4) that constitutes an upper layer from the focused node (Col. 4 lines 4-55 & Fig. 2), and

the lower-level N-gram language model (Col. 6 lines 11-20) generation and accumulation unit generates the lower-level N-gram language model for syntactic tree, using a second subtree (Col. 5 lines 42-63 & Fig. 4) that constitutes a lower layer from the focused node (Col. 4 lines 4-55 & Fig. 2)

However, Rigazio fails to teach morphemic analysis

Deligne teaches that the N-gram class model is defined as a language model that approximates a word N-gram in combinations of occurrence distributions of word-class N-grams and class-based words as shown by the following equation (this equation becomes equivalent to an HMM equation in morphological or morphemic analysis if word classes are replaced by parts of speech (Deligne Col. 18 lines 1-16).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Rigazio to incorporate morphemic analysis as taught by Deligne to allow for a multidimensional probabilistic method of prediction used to classify and model speech, wherein analysis can be performed on the smallest text units (i.e. morphemes) (Deligne Col. 18 lines 1-16).

Further, Ragazio fails to teach the higher-level N-gram language model of claim 1.

Deligne teaches well known limitations of previous technology, wherein Deligne teaches class versions of phrase based models can be defined in a way similar to the

way class version of N-gram models are defined, i.e., by assigning class labels to the phrases. In prior art it consists in first assigning word class labels to the words, and in then defining a phrase class label for each distinct phrase of word class labels. A drawback of this approach is that only phrases of the same length can be assigned the same class label. For example, the phrases "thank you" and "thank you very much" cannot be assigned the same class label, because being of different lengths, they will lead to different sequences of word class labels (Deligne Col. 2 lines 10-20).

Further, Deligne improves these limitations by teaching the clustering (classification process) of the variable-length phrases is explained. Recently, class-phrase based models have gained some attention, but usually like in Prior Art Reference 1, it assumes a previous clustering of the words. Typically, each word is first assigned a word-class label C.sub.k, then variable-length phrases, wherein the phrases "thank you for" and "thank you very much for" cannot be assigned the same class label. In the present preferred embodiment, it is proposed to address this limitation by directly clustering phrases instead of words (Deligne Col. 10 lines 43-60)

Furthermore, Deligne teaches the step ensures that the class assignment based on the mutual information criterion is optimal with respect to the current phrase distribution, and the step SS2 ensures that the bigram distribution of the phrases optimizes the likelihood calculated according to Equation (19) with the current class distribution. The training data are thus iteratively structured at a both paradigmatic and syntagmatic level in a fully integrated way (the terms paradigmatic and syntagmatic are both linguistic terms). That is, the paradigmatic relations between the phrases

expressed by the class assignment influence the reestimation of the bigram distribution of the phrases, while the bigram distribution of the phrases determines the subsequent class assignment (Deligne Col. 11 lines 29-43).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Rigazio to incorporate higher level language modeling as taught by Deligne to allow for optimal class assignment to account for sentence and word based modeling in speech recognition (Deligne Col. 10 lines 43-60).

Re claim 31, Ragazio teaches the language model generation and accumulation apparatus according to claim 1,

wherein the lower-level N-gram language model (Col. 6 lines 11-20) generation and accumulation unit is operable to represent a first sequence of words having a common linguistic property (Fig. 1 features) as the word string class, to generate and to accumulate, for each word string class, the lower-level N-gram language model that is obtained by modeling the first sequence of words included in the word string class (Col. 4 lines 30-55 & Fig. 2); and

the lower-level N-gram language model generation and accumulation unit is operable to generate and accumulate, for each word string class, the first sequence of words having the linguistic property (Fig. 1 features) indicated by the word string class (Col. 4 lines 30-55 $ Fig. 2).

However, Ragazio fails to teach each word included in the first sequence of words and each word included in the second sequence of words are respectively morphemes which are smallest linguistic units that have meaning

replace the first sequence of words modeled in the lower-level N-grams language model included in a text which is the sequence of words with a word string class corresponding to the first sequence of word

Deligne teaches that the N-gram class model is defined as a language model that approximates a word N-gram in combinations of occurrence distributions of word-class N-grams and class-based words as shown by the following equation (this equation becomes equivalent to an HMM equation in morphological or morphemic analysis if word classes are replaced by parts of speech (Deligne Col. 18 lines 1-16).

the higher-level N-gram language model generation and accumulation unit is operable to replace the first sequence of words modeled in the lower-level N-grams language model included in a text which is the sequence of words with a word string class corresponding to the first sequence of word, and to generate and to accumulate a higher-lever N-gram language model that is obtained by modeling the text which is the character string as a sequence of words that includes the word string class and a second sequence of words

Deligne also teaches well known limitations of previous technology, wherein Deligne teaches class versions of phrase based models can be defined in a way similar to the way class version of N-gram models are defined, i.e., by assigning class labels to

the phrases. In prior art it consists in first assigning word class labels to the words, and in then defining a phrase class label for each distinct phrase of word class labels. A drawback of this approach is that only phrases of the same length can be assigned the same class label. For example, the phrases "thank you" and "thank you very much" cannot be assigned the same class label, because being of different lengths, they will lead to different sequences of word class labels (Deligne Col. 2 lines 10-20).

Further, Deligne improves these limitations by teaching the clustering (classification process) of the variable-length phrases is explained. Recently, class-phrase based models have gained some attention, but usually like in Prior Art Reference 1, it assumes a previous clustering of the words. Typically, each word is first assigned a word-class label C.sub.k, then variable-length phrases, wherein the phrases "thank you for" and "thank you very much for" cannot be assigned the same class label. In the present preferred embodiment, it is proposed to address this limitation by directly clustering phrases instead of words (Deligne Col. 10 lines 43-60)

Furthermore, Deligne teaches the step ensures that the class assignment based on the mutual information criterion is optimal with respect to the current phrase distribution, and the step SS2 ensures that the bigram distribution of the phrases optimizes the likelihood calculated according to Equation (19) with the current class distribution. The training data are thus iteratively structured at a both paradigmatic and syntagmatic level in a fully integrated way (the terms paradigmatic and syntagmatic are both linguistic terms). That is, the paradigmatic relations between the phrases expressed by the class assignment influence the reestimation of the bigram distribution

of the phrases, while the bigram distribution of the phrases determines the subsequent class assignment (Deligne Col. 11 lines 29-43).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Rigazio to incorporate each word included in the first sequence of words and each word included in the second sequence of words are respectively morphemes which are smallest linguistic units that have meaning, replace the first sequence of words modeled in the lower-level N-grams language model included in a text which is the sequence of words with a word string class corresponding to the first sequence of word , and the higher-level N-gram language model generation and accumulation unit is operable to replace the first sequence of words modeled in the lower-level N-grams language model included in a text which is the sequence of words with a word string class corresponding to the first sequence of word, and to generate and to accumulate a higher-lever N-gram language model that is obtained by modeling the text which is the character string as a sequence of words that includes the word string class and a second sequence of words as taught by Deligne to allow for a multidimensional probabilistic method of prediction used to classify and model speech, wherein analysis can be performed on the smallest text units (i.e. morphemes) (Deligne Col. 18 lines 1-16 as well as optimal class assignment to account for sentence and word based modeling in speech recognition (Deligne Col. 10 lines 43-60).

Re claims 32-35, Rigazio teaches the speech recognition apparatus according to Claim 14, wherein, in the speech recognized from an input speech,

an alignment of words is recognized from a input speech, by referring to a recognition dictionary which describes pronunciation of the words (Col. 7 line 20 – Col. 8 line 20),

a sequence of words including the word string class is assumed in the alignment of words (Col. 4 lines 4-55 & Fig. 2),

However, Rigazio fails to teach the input speech is recognized based on (i) a probability that the words including the word string class appear in an order of appearance in the assumed sequence of words and (ii) a probability of an appearance of the words or the virtual word denoting the end of the word string class in an order of appearance in the word string class

Deligne teaches that the N-gram class model is defined as a language model that approximates a word N-gram in combinations of occurrence distributions of word-class N-grams and class-based words as shown by the following equation (this equation becomes equivalent to an HMM equation in morphological or morphemic analysis if word classes are replaced by parts of speech (Deligne Col. 18 lines 1-16).

Deligne also teaches well known limitations of previous technology, wherein Deligne teaches class versions of phrase based models can be defined in a way similar to the way class version of N-gram models are defined, i.e., by assigning class labels to the phrases. In prior art it consists in first assigning word class labels to the words, and in then defining a phrase class label for each distinct phrase of word class labels. A drawback of this approach is that only phrases of the same length can be assigned the same class label. For example, the phrases "thank you" and "thank you very much"

cannot be assigned the same class label, because being of different lengths, they will lead to different sequences of word class labels (Deligne Col. 2 lines 10-20).

Further, Deligne improves these limitations by teaching the clustering (classification process) of the variable-length phrases is explained. Recently, class-phrase based models have gained some attention, but usually like in Prior Art Reference 1, it assumes a previous clustering of the words. Typically, each word is first assigned a word-class label C.sub.k, then variable-length phrases, wherein the phrases "thank you for" and "thank you very much for" cannot be assigned the same class label. In the present preferred embodiment, it is proposed to address this limitation by directly clustering phrases instead of words (Deligne Col. 10 lines 43-60)

Furthermore, Deligne teaches the step ensures that the class assignment based on the mutual information criterion is optimal with respect to the current phrase distribution, and the step SS2 ensures that the bigram distribution of the phrases optimizes the likelihood calculated according to Equation (19) with the current class distribution. The training data are thus iteratively structured at a both paradigmatic and syntagmatic level in a fully integrated way (the terms paradigmatic and syntagmatic are both linguistic terms). That is, the paradigmatic relations between the phrases expressed by the class assignment influence the reestimation of the bigram distribution of the phrases, while the bigram distribution of the phrases determines the subsequent class assignment (Deligne Col. 11 lines 29-43).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Rigazio to incorporate the input speech is

recognized based on (i) a probability that the words including the word string class appear in an order of appearance in the assumed sequence of words as taught by Deligne to allow for optimal class assignment to account for sentence and word based modeling in speech recognition (Deligne Col. 10 lines 43-60).

However, Deligne in view of Rigazio fails to teach the virtual word denoting the end of the word string class in an order of appearance in the word string class

Millett teaches that it is necessary to increase the current `virtual word number` up to the beginning of the next word cluster boundary whenever the end of a data item is reached in the word stream. The only way to know this is to place a marker in the word stream signaling the end of a data item. For non-word level indexes, granules naturally fall within data items, so there is not a problem with a row in the granule cross reference table referring to more than one data item (Millett Col. 11 lines 19-29).

Further, Millett teaches that granule boundary markers 58 are used to demarcate the beginning and end of granules (e.g., "<MB>" for the beginning of a granule and "<ME>" for the end of a granule 60), as shown in FIG. 2. As used herein, the term "granule" and its derivatives refers to a predetermined set of text, or an indexing unit. The granule size determines the degree to which the location of a word within a document can be determined. For example, a document level granularity would be able to identify the document in which a word appears but not the page or paragraph. A paragraph level granularity would-be able to more precisely identify the paragraph within a document where a word appears, while a word level granularity would be able

to identify the sequential word location of a word (e.g., the first word of the document, the second word of the document, etc.). As the granularity increases and approaches word level granularity, the size and complexity of an index increases, but word locations can be more precisely defined. The purpose of the word stream 44 is to track the granules in which a word occurs, not the total number of occurrences of the word. (Millett Col. 4 line 50 – Col. 5 line 15).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Rigazio in view of Deligne to incorporate the virtual word denoting the end of the word string class in an order of appearance in the word string class as taught by Millett to allow for the identification of varying text (i.e. phrases or words or paragraphs), wherein markers (i.e. virtual words) are used to tag the beginning and end of the specified granule (i.e. phrases, words, paragraphs, etc.) (Millett Col. 4 line 50 – Col. 5 line 15).

4. Claims 6, 8, 10-12, 19-21, and 23-25 rejected under 35 U.S.C. 103(a) as being unpatentable over Rigazio et al. US 6182039 B1 (hereinafter Rigazio) in view of Deligne et al. US 6314399 B1 (hereinafter Deligne) and Millett et al. US 6584458 B1 (hereinafter Millett) and further in view of Hwang et al. US 20020082831 A1 (hereinafter Hwang).

Re claims 6 and 19, Rigazio teaches the language model generation and accumulation apparatus according to Claim 1,

wherein the lower-level N-gram language model (Col. 6 lines 11-20) generation and accumulation unit includes an exception word judgment unit operable to judge whether or not a specific word out of a plurality of words that appear in the word string class should be treated as an exception word (Col. 4 lines 4-55 & Fig. 2), based on a linguistic property of the specific word, and divides the exception word into (i) a syllable that is a basic phonetic unit constituting a pronunciation of the exception word (Col. 4 lines 4-55 & Fig. 2) and (ii) a unit that is obtained by combining syllables based on a judgment result the exception word being (Col. 4 lines 4-55 & Fig. 2),

the language model generation and accumulation apparatus further comprises a class dependent syllable N-gram generation and accumulation unit operable to generate class dependent syllable N-grams by modeling a sequence made up of the syllable and the unit obtained by combining syllables and by providing a language likelihood (Col. 1 lines 31-39) to the sequence in dependency on either the word string class or the linguistic property of the exception word (Col. 4 lines 4-55 & Fig. 2),

However, Ragazio fails to teach a higher-level N-gram language model

the language likelihood being a logarithm value of a probability.

Deligne teaches well known limitations of previous technology, wherein Deligne teaches class versions of phrase based models can be defined in a way similar to the way class version of N-gram models are defined, i.e., by assigning class labels to the phrases. In prior art it consists in first assigning word class labels to the words, and in then defining a phrase class label for each distinct phrase of word class labels. A drawback of this approach is that only phrases of the same length can be assigned the

same class label. For example, the phrases "thank you" and "thank you very much" cannot be assigned the same class label, because being of different lengths, they will lead to different sequences of word class labels (Deligne Col. 2 lines 10-20).

Further, Deligne improves these limitations by teaching the clustering (classification process) of the variable-length phrases is explained. Recently, class-phrase based models have gained some attention, but usually like in Prior Art Reference 1, it assumes a previous clustering of the words. Typically, each word is first assigned a word-class label C.sub.k, then variable-length phrases, wherein the phrases "thank you for" and "thank you very much for" cannot be assigned the same class label. In the present preferred embodiment, it is proposed to address this limitation by directly clustering phrases instead of words (Deligne Col. 10 lines 43-60)

Furthermore, Deligne teaches the step ensures that the class assignment based on the mutual information criterion is optimal with respect to the current phrase distribution, and the step SS2 ensures that the bigram distribution of the phrases optimizes the likelihood calculated according to Equation (19) with the current class distribution. The training data are thus iteratively structured at a both paradigmatic and syntagmatic level in a fully integrated way (the terms paradigmatic and syntagmatic are both linguistic terms). That is, the paradigmatic relations between the phrases expressed by the class assignment influence the reestimation of the bigram distribution of the phrases, while the bigram distribution of the phrases determines the subsequent class assignment (Deligne Col. 11 lines 29-43).

Additionally, Deligne teaches the use of a logarithmic probability in relation to n-gram word classification (Deligne Col. 18 lines 25-40)

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Rigazio to incorporate a higher-level N-gram language model and a language likelihood being a logarithm value of a probability as taught by Deligne to allow for a well known probabilistic method of prediction used to classify and model speech, wherein analysis can be performed on the smallest text units (i.e. morphemes) (Deligne Col. 18 lines 1-16).

However, Rigazio in view of Deligne and Millett fail to teach a word not being included as a constituent word of the word string class accumulate the generated class dependent syllable N-grams

Hwang teaches n-gram analysis of text as well as syllables (well known in the art to be non-morphemic, non-word, non-sentence, etc.), wherein Hwang teaches that each syllable-like unit is found in SLU language model 512, which in many embodiments is a trigram language model. Under one embodiment, each syllable-like unit in language model 512 is named such that the name describes all of the phonetic units that make up the syllable-like unit. Using this naming strategy, SLU engine 510 is able to identify the phonetic units associated with each syllable-like unit simply by examining the name associated with the syllable-like unit. For example, the syllable-like unit named EH_K_S, which is the first syllable in the word "exclamation", contains the phonemes EH, K and S (Hwang [0064]).

Further, Hwang teaches SLU engine 510 updates the score for a hypothesized sequence of syllable-like units by adding the language model score and acoustic model score of the next syllable-like unit to the sequence score. SLU engine 510 calculates the language model score based on the model score stored in SLU language model 512 for the next syllable-like unit to be added to the hypothesized sequence. In one embodiment, SLU language model 512 is a trigram model, and the model score is based on the next syllable-like unit and the last two syllable-like units in the sequence of units (Hwang [0066]).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Rigazio in view of Deligne and Millett to incorporate a word not being included as a constituent word of the word string class accumulate the generated class dependent syllable N-grams as taught by Hwang to allow for the proper identification of non-textual units such as syllables, wherein modeling can be phonetically implemented after progressing from paragraph to morpheme to syllable to find the combination/sequence of syllable that form an overall textual element located within text (Hwang [0064]).

Re claims 8, 10, and 23, Rigazio teaches the language model (Col. 6 lines 11-20) generation and accumulation apparatus according to Claim 7,

wherein the lower-level N-gram language model (Col. 6 lines 11-20) generation and accumulation unit includes

a language model generation exception word judgment unit operable to judge a specific word appearing in the second subtree (Col. 5 lines 42-63)as an exception word based on a predetermined linguistic property (Col. 4 lines 30-55 $ Fig. 2), the exception word being a word not being included as a constituent word of any subtree (Col. 4 lines 30-55 $ Fig. 2),

the lower-level N-gram language model generation and accumulation unit generates the lower-level N-gram language model (Col. 4 lines 30-55 $ Fig. 2)by dividing the exception word into (i) a syllable that is a basic phonetic unit constituting a pronunciation of the word (Col. 4 lines 30-55 $ Fig. 2) and (ii) a unit that is obtained by combining syllables, and then by modeling a sequence made up of the syllable and the unit obtained by combining syllables in dependency on a location of the exception word in the syntactic tree (Col. 5 lines 42-63) and on the linguistic property of the exception word (Col. 4 lines 30-55 $ Fig. 2)

However, Rigazio in view of Deligne and Millett fail to teach a word not being included as a constituent word of the word string class accumulate the generated class dependent syllable N-grams

Hwang teaches n-gram analysis of text as well as syllables (well known in the art to be non-morphemic, non-word, non-sentence, etc.), wherein Hwang teaches that each syllable-like unit is found in SLU language model 512, which in many embodiments is a trigram language model. Under one embodiment, each syllable-like unit in language model 512 is named such that the name describes all of the phonetic units that make up the syllable-like unit. Using this naming strategy, SLU engine 510 is able to identify the

phonetic units associated with each syllable-like unit simply by examining the name associated with the syllable-like unit. For example, the syllable-like unit named EH_K_S, which is the first syllable in the word "exclamation", contains the phonemes EH, K and S (Hwang [0064]).

Further, Hwang teaches SLU engine 510 updates the score for a hypothesized sequence of syllable-like units by adding the language model score and acoustic model score of the next syllable-like unit to the sequence score. SLU engine 510 calculates the language model score based on the model score stored in SLU language model 512 for the next syllable-like unit to be added to the hypothesized sequence. In one embodiment, SLU language model 512 is a trigram model, and the model score is based on the next syllable-like unit and the last two syllable-like units in the sequence of units (Hwang [0066]).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Rigazio in view of Deligne and Millett to incorporate dividing the exception word into (i) a syllable that is a basic phonetic unit constituting a pronunciation of the word and (ii) a unit that is obtained by combining syllables, and then by modeling a sequence made up of the syllable and the unit obtained by combining syllables in dependency on a location of the exception word as taught by Hwang to allow for the proper identification of non-textual units such as syllables, wherein modeling can be phonetically implemented after progressing from paragraph to morpheme to syllable to find the combination/sequence of syllable that form an overall textual element located within text (Hwang [0064]).

Re claims 11 and 12, Rigazio teaches the language model generation and accumulation apparatus according to Claim 1,

wherein the higher-level N-gram language model (Col. 6 lines 11-20) generation and accumulation unit generates the higher-level N-gram language model in which each (Col. 4 lines 30-55 $ Fig. 2)

However, Rigazio fails to teach a sequence of N words including the word string class is associated a probability at which said each sequence of N words

analyzing the first sequence of words within the word string class into one or more morphemes that are the smallest language units having meanings.

Deligne teaches well known limitations of previous technology, wherein Deligne teaches class versions of phrase based models can be defined in a way similar to the way class version of N-gram models are defined, i.e., by assigning class labels to the phrases. In prior art it consists in first assigning word class labels to the words, and in then defining a phrase class label for each distinct phrase of word class labels. A drawback of this approach is that only phrases of the same length can be assigned the same class label. For example, the phrases "thank you" and "thank you very much" cannot be assigned the same class label, because being of different lengths, they will lead to different sequences of word class labels (Deligne Col. 2 lines 10-20).

Further, Deligne improves these limitations by teaching the clustering (classification process) of the variable-length phrases is explained. Recently, class-

phrase based models have gained some attention, but usually like in Prior Art Reference 1, it assumes a previous clustering of the words. Typically, each word is first assigned a word-class label C.sub.k, then variable-length phrases, wherein the phrases "thank you for" and "thank you very much for" cannot be assigned the same class label. In the present preferred embodiment, it is proposed to address this limitation by directly clustering phrases instead of words (Deligne Col. 10 lines 43-60)

Furthermore, Deligne teaches the step ensures that the class assignment based on the mutual information criterion is optimal with respect to the current phrase distribution, and the step SS2 ensures that the bigram distribution of the phrases optimizes the likelihood calculated according to Equation (19) with the current class distribution. The training data are thus iteratively structured at a both paradigmatic and syntagmatic level in a fully integrated way (the terms paradigmatic and syntagmatic are both linguistic terms). That is, the paradigmatic relations between the phrases expressed by the class assignment influence the reestimation of the bigram distribution of the phrases, while the bigram distribution of the phrases determines the subsequent class assignment (Deligne Col. 11 lines 29-43).

Additionally, Deligne teaches the use of a logarithmic probability in relation to n-gram word classification (Deligne Col. 18 lines 25-40)

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Rigazio to incorporate a sequence of N words including the word string class is associated a probability at which said each sequence of N words analyzing the first sequence of words within the word string class

into one or more morphemes that are the smallest language units having meanings as taught by Deligne to allow for a well known probabilistic method of prediction used to classify and model speech, wherein analysis can be performed on the smallest text units (i.e. morphemes) (Deligne Col. 18 lines 1-16).

Re claim 20, Rigazio teaches the language model generation and accumulation apparatus according to Claim 19, further comprising

a syntactic tree generation unit operable to perform morphemic analysis as well as syntactic analysis of a text (Col. 5 lines 42-63), and generate a syntactic tree in which said-the text is structured by a plurality of layers, focusing on a node that is on said the syntactic tree (Col. 5 lines 42-63) and that has been selected on the basis of a predetermined criterion (Col. 4 lines 4-55 & Fig. 2),

wherein the higher-level N-gram language model (Col. 6 lines 11-20) generation and accumulation unit generates the higher-level N-gram language model for syntactic tree, using a first subtree (Col. 5 lines 42-63 & Fig. 4) that constitutes an upper layer from the focused node (Col. 4 lines 4-55 & Fig. 2), and

the lower-level N-gram language model (Col. 6 lines 11-20) generation and accumulation unit generates the lower-level N-gram language model for syntactic tree, using a second subtree (Col. 5 lines 42-63 & Fig. 4) that constitutes a lower layer from the focused node (Col. 4 lines 4-55 & Fig. 2)

the speech recognition apparatus comprises:

an acoustic processing unit operable to generate feature parameters from the speech (Col. 4 lines 30-55 $ Fig. 2);

a word comparison unit operable to compare a pronunciation of each word with each of the feature parameters (Col. 4 lines 30-55 $ Fig. 2), and generate a set of word hypotheses including an utterance segment of each word and an acoustic likelihood of each word (Col. 1 lines 31-39);

a word string hypothesis (Col. 12 lines 23-41) generation unit operable to generate a word string hypothesis from the set of word hypotheses with reference to the higher-level N-gram language model for syntactic tree (Col. 5 lines 42-63) and the lower-level N-gram language model for syntactic tree (Col. 5 lines 42-63), and generate a result of the speech recognition

However, Rigazio fails to teach a higher level n-gram modeling and morphemic analysis

Deligne teaches that the N-gram class model is defined as a language model that approximates a word N-gram in combinations of occurrence distributions of word-class N-grams and class-based words as shown by the following equation (this equation becomes equivalent to an HMM equation in morphological or morphemic analysis if word classes are replaced by parts of speech (Deligne Col. 18 lines 1-16).

Deligne also teaches well known limitations of previous technology, wherein Deligne teaches class versions of phrase based models can be defined in a way similar to the way class version of N-gram models are defined, i.e., by assigning class labels to the phrases. In prior art it consists in first assigning word class labels to the words, and

in then defining a phrase class label for each distinct phrase of word class labels. A drawback of this approach is that only phrases of the same length can be assigned the same class label. For example, the phrases "thank you" and "thank you very much" cannot be assigned the same class label, because being of different lengths, they will lead to different sequences of word class labels (Deligne Col. 2 lines 10-20).

Further, Deligne improves these limitations by teaching the clustering (classification process) of the variable-length phrases is explained. Recently, class-phrase based models have gained some attention, but usually like in Prior Art Reference 1, it assumes a previous clustering of the words. Typically, each word is first assigned a word-class label C.sub.k, then variable-length phrases, wherein the phrases "thank you for" and "thank you very much for" cannot be assigned the same class label. In the present preferred embodiment, it is proposed to address this limitation by directly clustering phrases instead of words (Deligne Col. 10 lines 43-60)

Furthermore, Deligne teaches the step ensures that the class assignment based on the mutual information criterion is optimal with respect to the current phrase distribution, and the step SS2 ensures that the bigram distribution of the phrases optimizes the likelihood calculated according to Equation (19) with the current class distribution. The training data are thus iteratively structured at a both paradigmatic and syntagmatic level in a fully integrated way (the terms paradigmatic and syntagmatic are both linguistic terms). That is, the paradigmatic relations between the phrases expressed by the class assignment influence the reestimation of the bigram distribution

of the phrases, while the bigram distribution of the phrases determines the subsequent class assignment (Deligne Col. 11 lines 29-43).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Rigazio to incorporate each word included in the first sequence of words and each word included in the second sequence of words are respectively morphemes which are smallest linguistic units that have meaning, replace the first sequence of words modeled in the lower-level N-grams language model included in a text which is the sequence of words with a word string class corresponding to the first sequence of word , and the higher-level N-gram language model generation and accumulation unit is operable to replace the first sequence of words modeled in the lower-level N-grams language model included in a text which is the sequence of words with a word string class corresponding to the first sequence of word, and to generate and to accumulate a higher-lever N-gram language model that is obtained by modeling the text which is the character string as a sequence of words that includes the word string class and a second sequence of words as taught by Deligne to allow for a multidimensional probabilistic method of prediction used to classify and model speech, wherein analysis can be performed on the smallest text units (i.e. morphemes) (Deligne Col. 18 lines 1-16 as well as optimal class assignment to account for sentence and word based modeling in speech recognition (Deligne Col. 10 lines 43-60).

Re claim 21, Rigazio teaches the apparatus according to Claim 20,

wherein the lower-level N-gram language model (Col. 6 lines 11-20) generation and accumulation unit includes

a language model generation exception word judgment unit operable to judge a specific word appearing in the second subtree (Col. 5 lines 42-63)as an exception word based on a predetermined linguistic property (Col. 4 lines 30-55 $ Fig. 2), the exception word being a word not being included as a constituent word of any subtree (Col. 4 lines 30-55 $ Fig. 2),

the lower-level N-gram language model generation and accumulation unit generates the lower-level N-gram language model (Col. 4 lines 30-55 $ Fig. 2)by dividing the exception word into (i) a syllable that is a basic phonetic unit constituting a pronunciation of the word (Col. 4 lines 30-55 $ Fig. 2) and (ii) a unit that is obtained by combining syllables, and then by modeling a sequence made up of the syllable and the unit obtained by combining syllables in dependency on a location of the exception word in the syntactic tree (Col. 5 lines 42-63) and on the linguistic property of the exception word (Col. 4 lines 30-55 $ Fig. 2)

the word string hypothesis generation unit generates the result of the speech recognition (Col. 12 lines 23-41).

However, Rigazio in view of Deligne and Millett fail to teach a word not being included as a constituent word of the word string class accumulate the generated class dependent syllable N-grams

Hwang teaches n-gram analysis of text as well as syllables (well known in the art to be non-morphemic, non-word, non-sentence, etc.), wherein Hwang teaches that each

syllable-like unit is found in SLU language model 512, which in many embodiments is a trigram language model. Under one embodiment, each syllable-like unit in language model 512 is named such that the name describes all of the phonetic units that make up the syllable-like unit. Using this naming strategy, SLU engine 510 is able to identify the phonetic units associated with each syllable-like unit simply by examining the name associated with the syllable-like unit. For example, the syllable-like unit named EH_K_S, which is the first syllable in the word "exclamation", contains the phonemes EH, K and S (Hwang [0064]).

Further, Hwang teaches SLU engine 510 updates the score for a hypothesized sequence of syllable-like units by adding the language model score and acoustic model score of the next syllable-like unit to the sequence score. SLU engine 510 calculates the language model score based on the model score stored in SLU language model 512 for the next syllable-like unit to be added to the hypothesized sequence. In one embodiment, SLU language model 512 is a trigram model, and the model score is based on the next syllable-like unit and the last two syllable-like units in the sequence of units (Hwang [0066]).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Rigazio in view of Deligne and Millett to incorporate a word not being included as a constituent word of the word string class accumulate the generated class dependent syllable N-grams as taught by Hwang to allow for the proper identification of non-textual units such as syllables, wherein modeling can be phonetically implemented after progressing from paragraph to

morpheme to syllable to find the combination/sequence of syllable that form an overall textual element located within text (Hwang [0064]).

Re claims 24 and 25, Rigazio teaches the speech recognition apparatus according to Claim 14,

wherein the higher-level N-gram language model (Col. 6 lines 11-20) generation and accumulation unit generates the higher-level N-gram language model in which each sequence of N words (Col. 4 lines 30-55 $ Fig. 2)

the speech recognition apparatus comprises

a word string hypothesis generation unit operable to evaluate a word string hypothesis (Col. 12 lines 23-41).

However, Ragazio fails to teach a higher-level N-gram language model

a word string class is associated with a probability at which the each sequence of words

multiplying each probability at which the each sequence of N words including the word string class occurs

Deligne teaches well known limitations of previous technology, wherein Deligne teaches class versions of phrase based models can be defined in a way similar to the way class version of N-gram models are defined, i.e., by assigning class labels to the phrases. In prior art it consists in first assigning word class labels to the words, and in then defining a phrase class label for each distinct phrase of word class labels. A drawback of this approach is that only phrases of the same length can be assigned the

same class label. For example, the phrases "thank you" and "thank you very much" cannot be assigned the same class label, because being of different lengths, they will lead to different sequences of word class labels (Deligne Col. 2 lines 10-20).

Further, Deligne improves these limitations by teaching the clustering (classification process) of the variable-length phrases is explained. Recently, class-phrase based models have gained some attention, but usually like in Prior Art Reference 1, it assumes a previous clustering of the words. Typically, each word is first assigned a word-class label C.sub.k, then variable-length phrases, wherein the phrases "thank you for" and "thank you very much for" cannot be assigned the same class label. In the present preferred embodiment, it is proposed to address this limitation by directly clustering phrases instead of words (Deligne Col. 10 lines 43-60)

Furthermore, Deligne teaches the step ensures that the class assignment based on the mutual information criterion is optimal with respect to the current phrase distribution, and the step SS2 ensures that the bigram distribution of the phrases optimizes the likelihood calculated according to Equation (19) with the current class distribution. The training data are thus iteratively structured at a both paradigmatic and syntagmatic level in a fully integrated way (the terms paradigmatic and syntagmatic are both linguistic terms). That is, the paradigmatic relations between the phrases expressed by the class assignment influence the reestimation of the bigram distribution of the phrases, while the bigram distribution of the phrases determines the subsequent class assignment (Deligne Col. 11 lines 29-43).

Additionally, Deligne teaches the use of a logarithmic probability in relation to n-gram word classification (Deligne Col. 18 lines 25-40)

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Rigazio to incorporate a higher-level N-gram language model, a word string class is associated with a probability at which the each sequence of words, multiplying each probability at which the each sequence of N words including the word string class occurs as taught by Deligne to allow for optimal probabilistic class assignment to account for sentence and word based modeling in speech recognition (Deligne Col. 10 lines 43-60).

Conclusion

5. **THIS ACTION IS MADE FINAL.** Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of

the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Michael C. Colucci whose telephone number is (571)-270-1847. The examiner can normally be reached on 9:30 am - 6:00 pm, Monday-Friday.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Richemond Dorvil can be reached on (571)-272-7602. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/Michael C Colucci/
Examiner, Art Unit 2626
Patent Examiner
AU 2626
(571)-270-1847
Examiner FAX: (571)-270-2847
Michael.Colucci@uspto.gov

/Vijay B. Chawan/
for Richemond Dorvil, SPE of Art Unit 2626